電腦軟體輔助編輯文字編——從《上博楚簡文字編》的編輯及其資料庫的建構談起

羅凡晸
國立台灣師範大學國文系助理教授

鄒濬智
國立台灣師範大學國文博士生、國立臺灣科技大學人文學科兼任講師

關鍵詞：

文字編，上海博館藏戰國楚竹書，字形資料庫，古文字研究

摘　要：

《上海博物館藏戰國楚竹書》自 2001 年 11 月正式公布第一冊以來，到現在（2007 年 5 月）為止已公布五冊，根據消息指出，第六冊將於 2007 年公布。學界投入上博文字研究者所在多有，然皆苦於沒有完整且編輯體例良好的文字編可供參詳取用。中央研究院歷史語言研究所副研究員袁國華先生有鑒於此，自上博楚竹書第一冊出版之後，即商請南台科大季旭昇教授，偕同筆者全力投入《上博文字編》及字形資料庫的建置及編輯工作。

為了加速文字編的編輯工作，袁先生商請筆者主持軟體開發設計與背景資料的建置，希望可以利用電腦的快速運算功能取代部份人力工作。到目前為止，本工作小組已完成上博一到五冊文字編的初步編輯，並將全部資料上傳網站試運轉。為了和學界分享軟體輔助編輯文字編及網站建置字形資料庫的心得，筆者特撰此文向學界報告我們的工作成果。

壹、前言

《上海博物館藏戰國楚竹書》（以下簡稱「上博簡」）自 2001 年 11 月正式公布第一冊以來，到現在（2007 年 5 月）爲止已公布五冊[1]，據傳，第六冊將於 2007 年公布[2]，屆時戰國楚竹書這批珍貴的文字材料將全部展現在世人眼前。

在進行古文字材料的研究過程當中，文字編的編纂堪稱首要之事，有了文字

[1] 陸續由上海古籍出版社出版。
[2] 參見萬卷樓圖書公司 2007 年 3 月 23 日公布之最新書目消息，http://www.wanjuan.com.tw/。

編，許多文字考釋及所延伸出來的研究工作將變得輕鬆許多。筆者爲了便於博士論文的寫作，曾初步編纂「上博楚竹書（一）文字編」[3]，另外吉林大學陳瓊《《上海博物館藏戰國楚竹書（一）》研究現況及文字編》（2005 年 4 月）及牛淑娟《《上海博物館藏戰國楚竹書（二）》研究現況及文字編》（2005 年 4 月）等也曾就上博竹書進行文字編的編纂，可惜的是，到目前爲止，因爲上楚竹書出版工作進行緩慢，完整出版一套「上博楚竹書文字編」可說是遙遙無期。如果我們先將已公布的上博（一）至上博（五）五冊內容進行文字編的編纂，並對 2007 年即將出版的第六冊進行相關準備工作，便利完整上博文字編及早出版，這對楚文字研究者來說是非常重要的。

有鑒於此，擁有《包山楚簡文字編》[4]、《郭店楚簡研究・第一卷・文字編》[5]、《望山楚簡校錄》[6]等書之豐富編輯經驗的中央研究院歷史語言研究所副研究員袁國華先生，商請擁有主編上博楚竹書系列讀本[7]經驗的南台科技大學季旭昇教授[8]，以及筆者進行《上博簡文字編》的編纂工作。疑難文字的隸定及考釋部分由袁先生及季教授等人負責，楚簡材料的基本處理及文字編後端的電子文字編建置則由筆者負責。[9]目前《上博簡文字編》已接近完成階段。

本文一在說明《上博簡文字編》編輯的過程中，筆者如何處理背景資料，以利便電腦軟體輔助文字編之編輯，二在說明電子版《上博簡文字編》資料庫後端設計的相關理念，三在比較傳統文字編編輯與電腦輔助編輯的優劣，文末並將評介二個與楚簡文字字形相關的網站資料庫。

貳、電子版《上博簡文字編》面臨的缺字問題及其解決之道

缺字問題到目前爲止依舊是古文字學界的一大夢魘，雖然 unicode 編碼方式理論上可以提供幾億個造字空間，看似可以解決這個惱人的問題，但是就目前國際編碼協定的執行速度來看，unicode 編碼對古文字學界的前沿研究來說可以說是緩不濟急。從資料庫的建置來看，《上博簡文字編》面臨的第一個問題便是缺字的處理。以下我們將分項說明本工作如何解決資料庫缺字問題。

一、利用構字式並輔以字形圖片顯示unicode缺字

[3] 附於羅凡晸：《古文字資料庫建再研究——以上海博物館藏戰國楚竹書(一)爲例》（台北：台師大國文系博士論文，2003 年）書末。
[4] 臺北：藝文印書館，1992 年。
[5] 臺北：藝文印書館，1999 年。
[6] 臺北：藝文印書館，2004 年。
[7] 自 2003 年起陸續由臺北萬卷樓出版。
[8] 前國立台灣師範大學國文學系教授。
[9] 本文有關電腦軟體輔助文字編編輯的理念說明及文末評介其他楚簡字形資料庫的部份，由羅凡晸負責撰寫；古文字字形資料處理方式及傳統、電腦輔助編輯文字編的比較部份，由鄒濬智負責撰寫。

中央研究院資訊所研發之「漢字構形資料庫」具有以下特色與功能：[10]

（一）、資料庫以漢字的字形結構式（以下簡稱構字式）取代傳統的交換碼，徹底解決漢字的缺字問題。

（二）、資料庫以漢字風格碼表現不同風格的同字異寫形態，比傳統字型名稱（譬如標楷體、粗隸體等等）更爲方便。

（三）、資料庫可檢索不同歷史時期的漢字字形及其構形分析。

（四）、資料庫介面提供各類字書索引，增加查詢字形出處的便利。

（五）、資料庫適用於所有微軟中文視窗軟體。

（六）、資料庫與現有的電腦應用程式相容無礙，安裝便利。

目前所知，該資料庫已收入滕壬生《楚系簡帛文字編》[11]約三千多字頭、上萬楚文字字形，對解決一般楚簡相關研究的電腦缺字問題很有助益。故本電子文字編編輯計畫原則上在安裝該資料庫的電腦中進行。

只是目前「漢字構形資料庫」資料量不足，無法立即解決上博簡一到五冊共654 個資料庫無法支援顯示缺字的問題。[12]爲了解決此一問題，我們從上博原書上掃瞄其簡文釋文，並使用Ulead Photoimpact等影像處理軟體一一將無法由「漢字構形資料庫」顯示的難字釋文切割存檔，並將此一造字圖片分別存成BMP檔及GIF檔，解析度均設爲 64*64——存成BMP是爲了將這些字形匯入到造字程式當中，做爲楚文字背景資料輸入時所使用；存成GIF檔是爲了將來要將所有資料上網時，讓這些難字字形得以順利顯示在網頁上。

二、利用Java Applet方式於網頁表達unicode缺字

在「漢字構形資料庫」中，爲了順利讓這些缺字能夠在網頁中順利呈現，「漢字構形資料庫」軟體提供了「缺字處理增益集」來轉換構字式，讓網頁得以顯示unicode缺字。[13]

參、上博楚簡圖版字形的處理及相關字形資料的建立

本工作小組原意將上博圖版數位化後編輯成文字編出版，以下我們將由「圖版取得方面」、「字形取得方面」、「字形資料建立」、「字形所處文例呈現的問題」等四個方面將圖版數位化的步驟做詳細的說明。

一、取得圖版電子檔

本工作使用德國「LEMAX」掃瞄器。此掃瞄器雖然影像取得的速度上比較慢，其圖版色彩品質十分穩定，且色差較小。掃瞄後的各版圖片以 600DPI 解析

---

[10] 鄒濬智、莊德明、袁國華：〈「楚系簡帛文字構形資料庫」的建置及其與「漢字構形資料庫」的整合〉，第十六屆中國文字學會，高雄：高雄師範大學國文系，2005 年 5 月。

[11] 滕壬生：《楚系簡帛文字編》，武漢：湖北教育出版社，1995 年。

[12] 根據本資料庫「上博楚竹書電子文字編」的初步統計，上博一至上博五共有 17,586 字。至於上博六截至完稿之日尚未公布，因此實際字數尚未得知。

[13] 中央研究院資訊所文獻處理實驗室、教育部圖書館自動化及網路化策略推動委員會：《中文電腦缺字解決方案：漢字構形資料庫使用手冊》，2002 年 7 月，頁 38-46。

度，全彩 BMP 格式存入資料夾，全簡圖檔命名方式爲「上博篇名＋半形二位數簡號」，譬如《上博一・緇衣》第三簡簡影圖片命名爲「緇衣 03」；存檔之資料夾則依上博楚簡冊數命名，如「上博一」、「上博二」等等，依此類推。

二、取得字形電子檔

本工作以 Ulead Phtoimpact 軟體一一開啓簡影。使用「選取」→「拖曳」方式將字形一一另存新檔，殘缺不全之字亦要切割存檔。因爲簡牘中的符號往往蘊藏有豐富的簡牘制度相關文化訊息，若所切割之字形在圖版簡影中連同重文、合文符等其他簡牘符號一齊出現，則將該字連同其相連之符號一併切下。

單一字形存檔格式爲全彩 JPG，解析度爲 600DPI；存檔之彩檔圖案命名方式爲「半形二位篇名流水序號-半形二位數簡號-半形二位數簡次」，其中「半形二位篇名流水序號」如下所示：

01	上博（一）：孔子詩論
02	上博（一）：緇衣
03	上博（一）：性情論
04	上博（二）：民之父母
05	上博（二）：子羔
06	上博（二）：魯邦大旱
07	上博（二）：從政甲篇
08	上博（二）：從政乙篇
09	上博（二）：昔者君老
10	上博（二）：容成氏
11	上博（三）：周易
12	上博（三）：中弓
13	上博（三）：亙先
14	上博（三）：彭祖
15	上博（四）：采風曲目
16	上博（四）：逸詩
17	上博（四）：昭王毀室與龔之準
18	上博（四）：柬大王泊旱
19	上博（四）：內禮
20	上博（四）：相邦之道
21	上博（四）：曹沫之陣
22	上博（五）：競建內之
23	上博（五）：鮑叔牙與隰朋之諫
24	上博（五）：季庚子問於孔子
25	上博（五）：姑成家父
26	上博（五）：君子爲禮

27	上博（五）：弟子問
28	上博（五）：三德
29	上博（五）：鬼神之明融師有成氏

屆時上博六正式公布，其篇名流水號將承接上表序號。根據這樣的圖版單字命名方式，如以《上博一・孔子詩論》第一簡第一個字爲例，它的圖版編號即爲「01-01-01」，依此類推。存檔之所處資料夾以篇名流水序號及中文篇名命名。如上述檔案即應存入名爲「（01）孔子詩論」資料夾之中。

三、建立字形資料

待所有上博楚簡字形全部切割完畢之後，我們使用 Microsoft Office Excel 軟體，建立：「流水號」、「字形編號」、「圖檔編號」、「嚴式隸定」、「寬式隸定」、「字根」、「倉頡碼」、「倉頡英文碼」、「注音碼」、「注音英文碼」、「備註」諸欄位。以下就需要說明之欄位建立原因及資料輸入內容分述如下：

（一）、流水號

待上博楚竹書全部出版，且所有字形全部切割並將各字形資料輸入此一 Excel 工作表單後，給予全部字形統一的流水號，以利資料庫管理。此流水號可由 Excel 內建功能自動產出。

（二）、圖檔編號

輸入各字形之圖檔編號，可供資料庫使用者抓取與之相對應之楚文字字形，亦裨利進一步文字編的編輯工作。此圖檔編號使用 ACDSee「工具」中的「建立檔案清單」功能，先匯出上博楚簡各篇簡文的字形檔案之檔名純文字檔，再剪貼至此欄位中。

（三）、嚴式隸定、寬式隸定

輸入字形之嚴式隸定，是爲了提供文字學研究者據嚴格隸定字形搜尋字形背景資料之用；輸入字形寬式隸定字形，是爲了提供初學者或其他學術領域之研究者據其一般之書寫習慣搜尋字形背景資料之用。嚴式及寬式隸定在初步輸入時依原書考釋。

（四）、字根

依嚴式隸定之形，拆至最小成意義偏旁者（字根）。字根欄之資料除供文字研究者知曉該楚文字形之構成資訊外，若資料庫使用者未安裝本計畫所提供之造字軟體，也可依所欲查尋之文字字根或偏旁部首搜尋其資料和出處。字根或有不詳及難以隸定者，於欄位中輸入「 ❓ 」代替。

（五）、倉頡碼、注音碼

若資料庫使用者未安裝本計畫所提供之造字軟體，也可依所欲查尋之文字嚴式隸定字形的倉頡碼或注音碼搜尋其資料和出處。原則上每一嚴式隸定字形皆可拆出倉頡碼，但有些楚文字讀音不明，未必能拆出注音碼。未能拆出字碼的欄位，全部輸入半形大寫「N」，以免資料庫在進行搜尋時發生程式上的錯亂。另外，此兩欄位之英文碼欄位則輸入各楚簡字形嚴式隸定後之拆字碼，以利使用不同中文輸入法之使用者參考。沒有拆字英文碼者，亦在欄位中輸入半形大寫「N」，以免資料庫軟體發生錯亂。

（六）、備註

於此欄輸入本文字編之參與學者對原書考釋成果的不同意見，包括對字音通讀的不同看法、對字形考釋的討論意見等等。此欄位之資料尚在陸續構建中。

四、字形所處文例呈現的問題

本工作將上博各冊之釋文採 Microsoft Office Access 軟體全面建檔並置於同一個資料表中。各簡之簡文在 Access 的資料表中，共設計二個欄位，第一個欄位輸入「上博篇名之流水序號+半形二位數簡號」，如《上博一・孔子詩論》第一簡的編號爲「01-01」；第二個欄位則輸入該簡上所有的釋文，然後逐簡逐條建檔。屆時再利用 ASP 動態網頁伺服器語言，配合 SQL 查詢語言，程式內部目前設計成顯示查詢字形的前後各一句話的釋文。

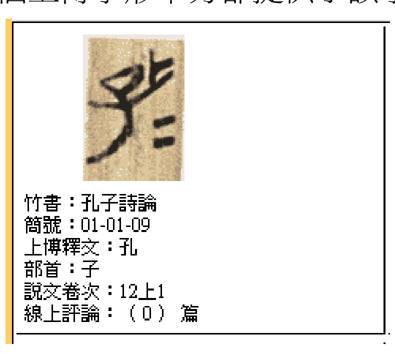
肆、《上博簡文字編》資料庫的設計

目前筆者所建置的《上博簡文字編》資料庫提供若干查詢功能：「單字查詢與 Metadata 的呈現」、「各篇竹書內容分頁查詢」、「字頭查詢」、「上博簡電子文字編查詢」等，以下將逐項說明之。

一、單字查詢與 **Metadata** 的呈現

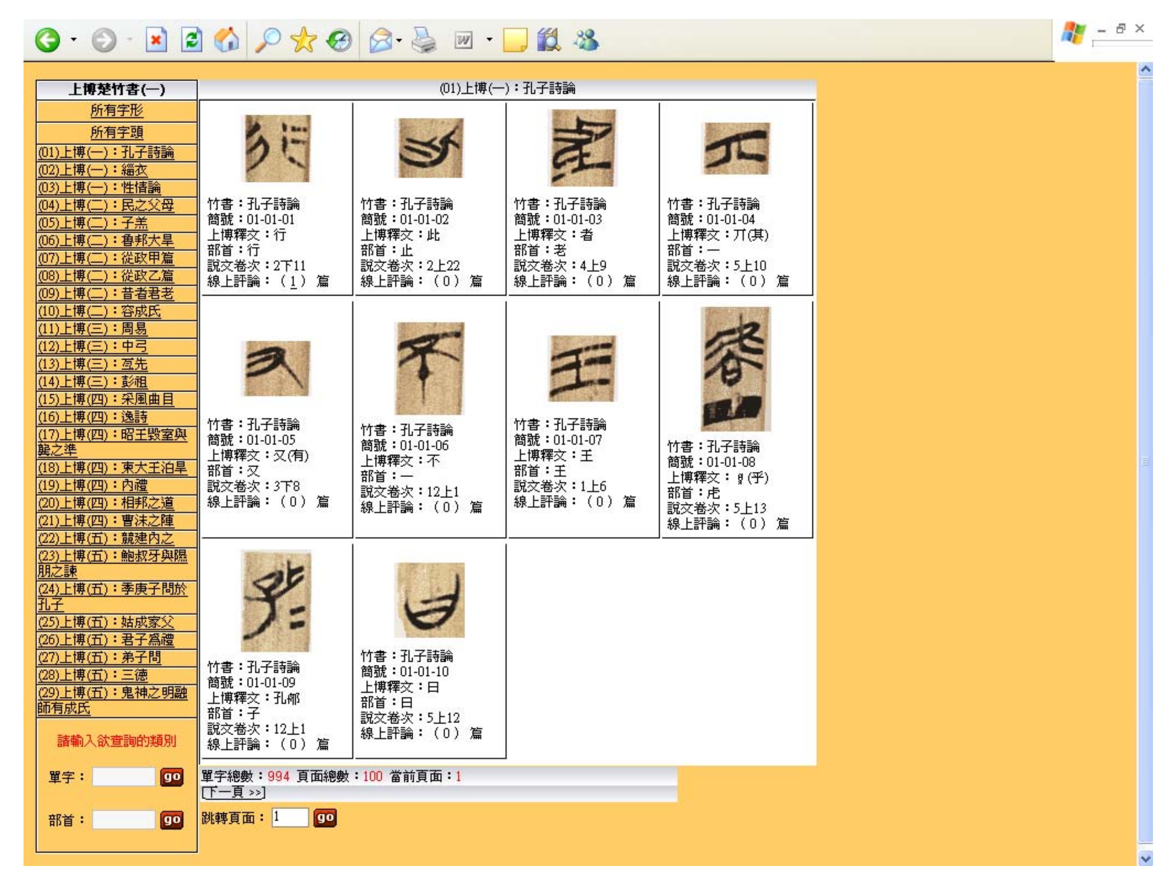


上圖所呈現的是單字搜尋的功能與單字圖版的相關 Metadata，以「孔」字爲例，本系統共找到了三十七筆資料（目前收錄上博一至上博五所有的字形圖片）。除了呈現楚簡字形外，每個上博字形下方都提供了該字的相關資訊：



第一項「竹書」：說明字形出處之篇名；第二項「簡號」：說明字形所在篇數、簡號及字序；第三項「上博釋文」：呈現上博原書之釋文隸定；第四項「部首」：以康熙字典部首分類方式——提供使用者多一種字形查詢方式的選擇；第五項「說文卷次」：便利使用者進一步翻查《說文》；第六項「線上評論」：如果使用者對此字有任何意見，可以在此處提問、留言，以達到網站互動教學及學習的效果。

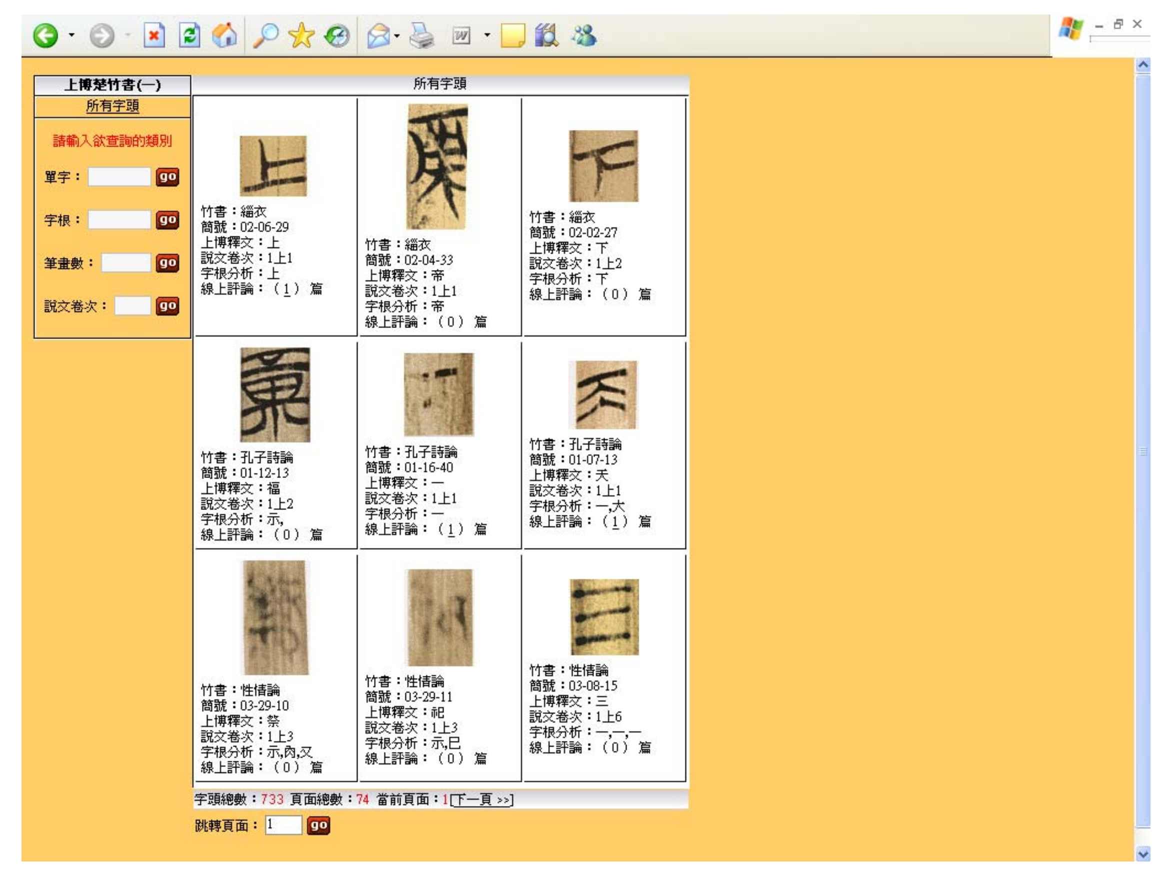
二、各篇竹書內容分頁查詢



上博一至上博五，根據原整理者所整理的結果，共有二十九篇。以查詢第一篇〈孔子詩論〉爲例，當我們在畫面左邊選擇「（01）上博（一）：孔子詩論」並按下滑鼠左鍵時：1.畫面右邊呈現「每頁顯示十筆資料」——於此使用者可依不同需求而調整每頁所呈現的資料數量；2.畫面下方則顯示〈孔子詩論〉共有 994 個單字等資料。這些資訊的提供有助於使用者了解該篇竹書的字數及篇章長度。

三、字頭查詢

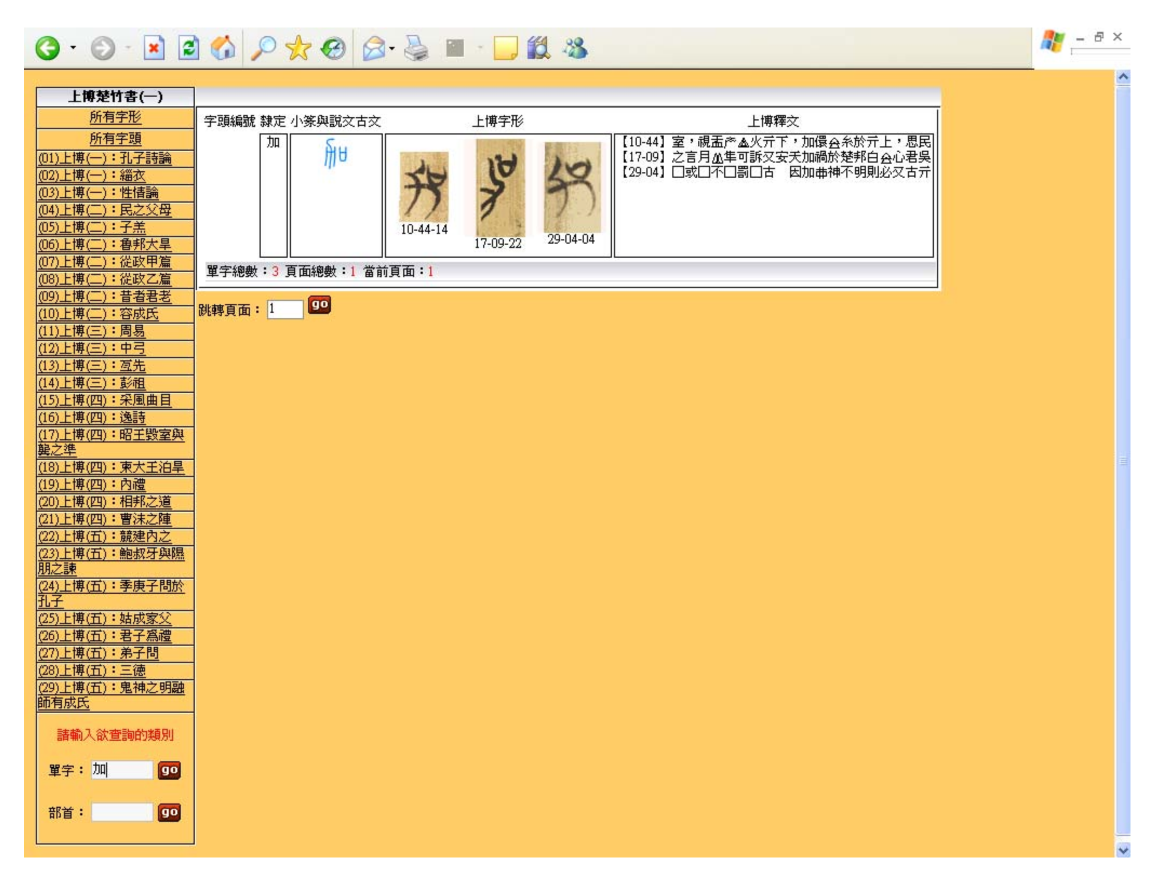
當我們按下畫面左方「所有字頭」，畫面如下所示：



這個查詢畫面能讓使用者了解上博一至上博五的一萬七千多個字裡共出現了多少不同的單字。目前爲止，包含已識之字與未識之字，本計畫已整理出約三千多個字頭。[14]

四、上博簡電子文字編查詢

[14]由於上博二至五的資料尚未完全上傳至網站，目前所呈現的字頭總數只是〈孔子詩論〉、〈緇衣〉、〈性情論〉等三篇楚竹書的字頭結果。



以「加」字爲例，當我們在網頁左下方「單字」的查詢欄裡輸入「加」字，網頁畫面的右邊出現查詢結果。查詢結果畫面共分五個欄位：第一個欄位是「字頭編號」，所呈現的是我們依軟體需求所編出的字頭編號[15]；第二個欄位是「隸定」，所呈現的是「寬式隸定」的結果；第三個欄位是「小篆與說文古文」，目前本計畫以北京師範大學所製作的說文小篆爲底稿，配合Ulead Photoimpact等影像處理軟體將小篆字形以圖版方式呈現；第四個欄位是「上博字形」，目前網站上陸續收有上博一至上博五的單字圖版資料，待未來上博六公布後，我們將即刻進行資料的統整匯入工作；第五個欄位是「上博釋文」，目前所呈現的是利用構字式方式呈現缺字字形的釋文，原則上只要了解「漢字構形資料庫」所使用之構字方式，即可了解該缺字之字形。目前資料庫所收釋文的內容，主要是依照上博原書的考釋成果，而學者們豐碩的後續文字考釋研究成果，屆時也將以不同的方式呈現出來。

伍、「軟體輔助編輯文字編」與「傳統人工編輯文字編」之比較

一、傳統文字編與軟體輔助編輯文字編之編排流程

[15] 由於資料表尚有部分未完成，所以此部份還未能將完整結果呈現出來。

（一）、傳統文字編編排流程

傳統文字編編排流程依取用字形的方式不同約有二種不同的編輯方法：

1.編輯者自行描摹文字字形：若編輯者書法功力深厚，多半會採取此種編輯方式，如張守中的《包山楚簡文字編》[16]及《郭店楚簡文字編》[17]、滕壬生的《楚系簡帛文字編》[18]等。若編輯者自行描摹，需事先準備抄錄字形之紙卡，以字形之部首或筆畫做爲紙卡之索引，而後逐頁翻閱簡帛之圖版，再一字一字連同字形之出處（文章篇號＋簡號）分別以毛筆錄入紙卡之中，待翻閱完整批簡帛圖版，復謄抄在已設計好「楷字字頭」、「字頭對應之小篆字形」、「簡帛字形」欄位之表格當中。[19]各字頭之排列次序依《說文》五百四十部首，若爲《說文》所無，則附於最爲形近的部首之後。最後依排序後之字頭做出文字編書後頁數索引後送印刷廠製版印刷。[20]

2.編輯者由簡帛圖版剪下簡帛字形存真：爲保有簡帛字形的寫實感，近來文字編編輯者多半從簡帛圖版上直接將簡帛字形剪下使用。剪下之字形紙片，亦依其部首或筆畫分類暫時貼在已製作好部首索引的紙卡上（字形出處亦在暫貼紙片時抄上），並同時先抄錄好字形相關的文例。待整批圖版字形全部剪裁分類後重新穩固黏貼到已設計好欄位的紙卡之中，方送印刷廠製版印刷。紙卡順序之安排及書後字頭索引之製作如前述第一種的編輯方式。

（二）、軟體輔助編輯文字編

人工掃瞄切割簡帛圖版字形，並將各個字形的基本資料建檔，同時建立簡文的釋文電子檔。爾後使用電腦軟體自動產出具有「楷字字頭」、「小篆字頭」、「字形及字形出處」、「文例」的內容頁，書後部首或筆畫索引亦使用電腦軟體產出。除部首或筆畫索引，並另附有注音碼及倉頡碼索引。

二、傳統文字編編輯與軟體輔助文字編編輯之比較

（一）、字編編輯時間上的比較

傳統文字編的編輯工作十分依賴人力，常耗掉從事者極龐大的時間，但若使用軟體輔助編輯的話，軟體會自動跑出成果，如此可省下煩雜瑣碎的編輯工作。又後者之前置作業，因配合軟體功能而建立有完整的作業標準流程。遵循標準流程，我們可以在同一個時間投入大量的學士級以上人力分頭進行不同的機械性工作，這大大的加速了文字編的出版。

---

[16] 張守中：《包山楚簡文字編》，北京：文物出版社，1996 年。
[17] 張守中：《郭店楚簡文字編》，北京：文物出版社，2000 年。
[18] 滕壬生：《楚系簡帛文字編》，武漢：湖北教育出版社，1995 年。
[19] 依各編輯者之不同編輯習慣，其所設計之欄位亦不盡相同。
[20] 索引多半分爲筆畫及部首索引二種。用力深者，在抄錄字形時會一併抄錄字形出處的上下文例，以供使用者參考。

以一個碩士級人員處理一本上博楚竹書爲例[21]，經估算：（一）高階掃瞄圖版（約一百五十頁左右）僅需一～二個工作天；（二）切割字形圖檔（約三千～四千字）僅需二～三個工作天；（三）爲各個字形圖檔建立Excel檔案資料僅需不到二十個工作天（包括標出待討論的疑難文字）；（四）二位以上的研究人員討論原考釋釋文中的疑難文字僅需十～十五個工作天。如此，由原始圖版到整理成有秩序的資料僅需不到一個月，加上研究人員討論出最終釋文版本的時間，只花不到一個半月。

（二）、字編編輯錯誤率上的比較

徒手抄錄簡帛字形，其中存在的第一個人爲干擾即「個人對字形筆畫有無、筆畫走向的判斷」。一般書法家多半從隸、楷、草、行書等下手習帖，較少有先從先秦古文字下手者；而書法家兼有古文字研究者身份者又不多，故徒手抄錄簡帛字形時，若抄錄者的古文字素養不夠，常會有抄錄字形失真的問題。

人畢竟是血肉之軀，長久工作，本易疲勞，疲勞便容易出錯。如此，在輾轉抄錄字形的過程當中也可能發生漏抄或錯抄情事，因而，其所編製的文字編，在字形準確度及寫真度上就會出現瑕疵。而錯抄或漏抄的情況，不止可能發生在文字字形的抄錄上，字形出處及書後索引的謄寫也可能會出現這樣的問題。

將字形從簡帛圖版上剪下來使用，此舉雖可避免字形抄錄失真的問題，但在整理暫時貼存的字形紙片時，極易遺失若干字形紙片，這也會造成其最終編輯成果無法全面反應該批簡帛圖版文字真貌的情況。而字形出處及書後索引的謄寫亦是訴諸人工，一樣無法完全避免誤抄或漏抄的情況發生。

然而使用軟體輔助編輯文字編，在輸入各個字形背景資料時雖有可能會有人工上的失誤，但因爲資料是配合電腦軟體所建置起來的，表格資料一個蘿蔔一個坑，從各冊到各篇到各簡到各字都編號依序建檔，於後續的工作當中極易從電子檔案上校對出人爲所造成的錯誤。

（三）、文字編資訊蘊涵強度的比較

傳統標誌字形出處的方式，泰半採用出處篇號＋簡號，若一簡有一字重複出現時，使用傳統文字編，使用者根本無法分辨何者在該簡中較先出現、何者較晚——傳統文字編無法反應字形於簡牘、簡帛中的出現字序。但若使用軟體輔助編，在建立字形 Excel 檔時必須全面建立了字形的所有資料，其所產出的文字編所提供的字序等其他資訊一定較傳統文字編多，十分便利學者進行同簡同字異形等其他比對研究。

此外，使用軟體輔助文字編時是採用高規格彩色掃瞄圖檔字形，依此操作所產出的紙本文字編，其上的古字形解晰度亦較傳統文字編爲佳，這對若干殘阞模糊字跡或簡牘符號的判斷極有助益。

[21] 每冊上博楚竹書約有三千餘字。

（四）、字編資料運用靈活度的比較

傳統文字編的使用，侷限在實體書面，研究或教學使用時，不論查詢的是字形或文例，使用者總是花了很多時間在翻查索引；而使用軟體輔助文字編，其Database隨時可以依不同使用需求建立一特定功能的字形資料庫，不論是要查詢單字字形、文例，或作多字字形、文例之間的比較，都遠較傳統文字編方便。

陸、與楚簡字形相關之網頁資料庫評介

除了本計畫正在努力建置的電子版《上博簡文字編》之外，筆者還發現若干已公布給學界使用的楚簡字形相關網頁資料庫，在此略作評介：

一、瀚堂典藏

（一）、資料庫內容基本介紹



「瀚堂典藏」的資料庫架構分別為「小學工具類資料」、「出土文獻類資料」、

「傳世文獻類資料」、「專題文獻類資料」等四大類。[22]其中，「出土文獻類資料」裡面前共收錄了「九店楚簡」五六號墓與六二一號墓、「武威漢簡」、「張家山漢簡」、「銀雀山漢簡」、「簡帛文」等，如下圖左方所示：



我們可以看到與「楚簡」有關的資料，目前收錄有「九店楚簡」五六號墓及六二一號墓、「簡帛文」二項類別。但「簡帛文」中到底收錄了哪些資料，網站的「說明」當中並未言及。

就目前所公布的楚簡圖版來看，《上海博物館藏戰國楚竹書》（五）一書最爲新近，其中有一「**頪**」字見於上博五「融師有成氏」第六簡，文例爲「**頪**獸非鼠」。爲了測試該網站所收資料究竟有無及時更新，本文以「**頪**」字查詢。結果我們在該系統查到 14 筆與「**頪**」相關的資料，如下所示：

---

[22] 「龍語瀚堂典籍資料庫——構建漢學研究及傳播的數位化新平臺」，http://www.hytung.cn/zh-tw/Introduction.aspx 。



以上十四個「頪」字的字形來源包含了上博二、上博一、郭店簡等資料。如以本文撰寫時（2007 年 3 月 10 日）的查詢結果來看，目前這個系統至少尚未將上博五的資料納入。

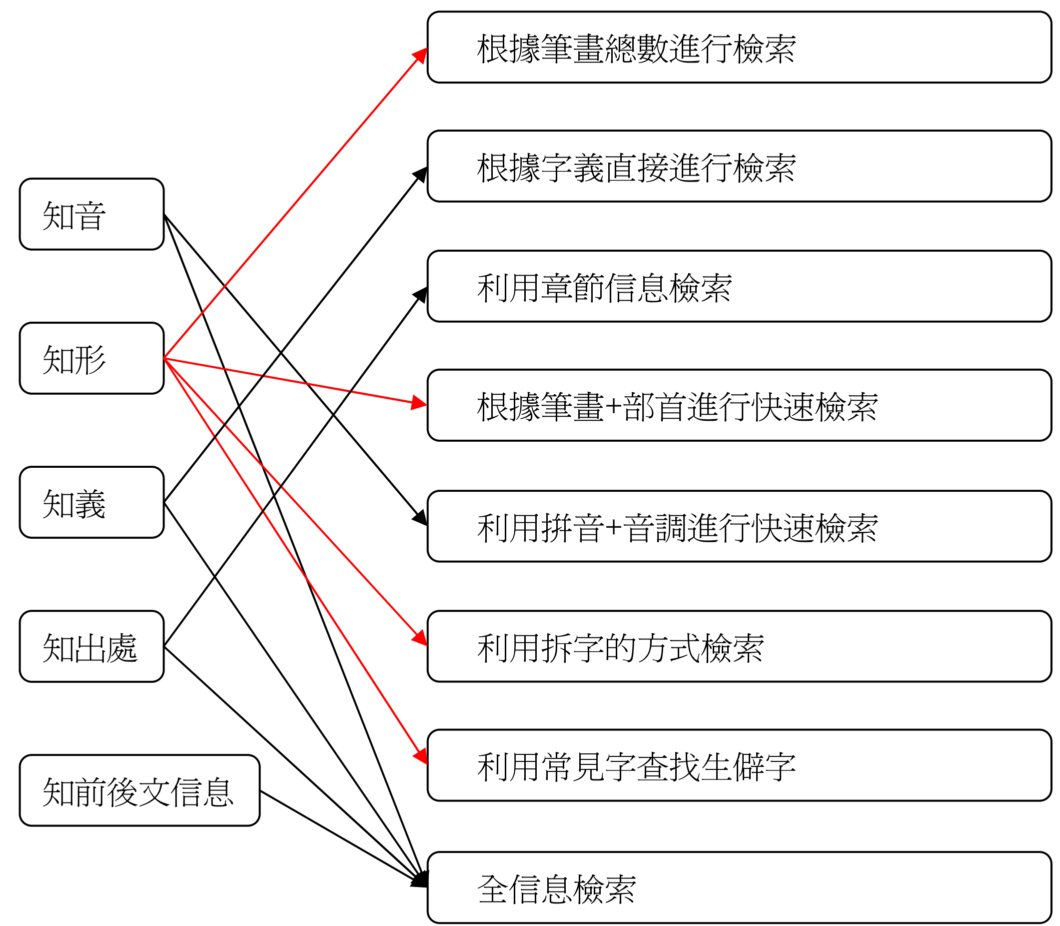
（二）、檢索方式介紹

根據網頁說明，「龍語瀚堂典籍書庫」的檢索方法主要有以下八種[23]：1.根據筆劃總數進行檢索；2.根據字義直接進行檢索；3.利用章節資訊檢索；4.根據「筆劃+部首」進行快速檢索；5.利用「拼音+音調」進行快速檢索；6.利用拆字的方式檢索；7.利用常見字查找生僻字；8.全資訊檢索。如以使用者所知的搜尋條件出發，則由「知音」、「知形」、「知義」、「知出處」、「知前後文資訊」等五種方式搜尋起。結合上述八種方法，該資料庫提供了許多查詢的便利性。其搜尋功能的設計理念如下圖所示[24]：

---

[23] 該網頁本身的分類共有八種，但其本身誤植爲「七」種。見 http://www.hytung.cn/mht/SearchHelp.htm。

[24] 改繪自：http://www.hytung.cn/mht/SearchHelp.htm。



如以查詢「簡帛」文字的角度思考，我們多半會以「知形」的方式進行查詢。以「知形」的查詢方法在該網站查詢，若使用者欲查詢罕見字時，網站的說明檔舉「岦」字爲例，使用者必須先在「unicode 字典」（該資料庫中的其中一部分）中找到相應的字，之後再複製這個字，轉貼到檢索框裡，才能夠勾選其他資料庫進行後續的查詢動作。

網站另外提供「根據筆畫總數進行檢索」、「根據筆畫+部首進行快速檢索」、「利用拆字的方式檢索」、「利用常見字查找生僻字」等四種功能。「筆畫總數」及「筆畫+部首」二種方式，使用者只要正確算出該字的「楷體」筆畫數，原則上可以搜尋到想要搜尋的字，不同之處是如以「筆畫總數」查詢，其呈現結果可能會跑出很多不相干的字；如以「筆畫+部首」的方式查詢，可稍稍縮小查詢結果。「利用常見字查找生僻字」，以「燊」字爲例，使用者只要輸入「燊」字上面所從的「焱」字，並且在檢索欄位中輸入，就可以找到「燊」字的相關資料。只是依此查詢方式，使用者仍會找到多筆不相關的字形，必須再多花時間自行判斷查詢結果是否就是自己所需的資料。

瀚堂網站所提供的資料查詢方法，其所呈現出來的結果，有很大部份可能是使用者不需要的資訊。雖然站在使用者的角度，氾濫的資訊強過資訊過少，但如果網站在設計上能提供更多面向的檢索方式，如Metadata的後設資料庫設計與建立[25]，或許可以提高使用者搜尋資料的效率與準確度。

[25] 讓使用者可以依據不同的字形條件，找到更直接且明確的資料。

二、戰國楚簡帛文字典型形體檢索系統

華東師範大學中國文字研究與應用中心建構一系列的古文字資源庫[26]，其中包含了「戰國楚簡帛文字典型形體檢索系統」[27]，根據該單位網頁的介紹內容，這個檢索系統是在《戰國楚文獻檢索系統》[28]的基礎上審核、增補完成的。「戰國楚簡帛文字典型形體檢索系統」所收錄的楚簡材料有：

著錄書名	編輯者	出版社及出版時間
《包山楚簡》	湖北省荊沙鐵路考古隊	文物出版社，1991
《長沙子彈庫戰國楚帛書研究》	李零	中華書局，1985
《郭店楚墓竹簡》	荊門市博物館	文物出版，1998
《九店楚簡》	湖北省文物考古研究所、北京大學中文系	中華書局，2000
《上博楚竹書》一	馬承源主編	上海古籍出版社，2000
《上博楚竹書》二	馬承源主編	上海古籍出版社，2002
《上博楚竹書》三	馬承源主編	上海古籍出版社，2004
《上博楚竹書》四	馬承源主編	上海古籍出版社，2005
《上博楚竹書》五	馬承源主編	上海古籍出版社，2006
《曾侯乙墓》	湖北省博物館編	文物出版社年，1989
《戰國楚竹簡彙編》	商承祚	齊魯書社，1995
《葛陵新蔡楚墓》	河南省文物考古研究所	大象出版社，2003

該資料庫的設計，是以「字」作爲基本檢索單位，網頁的說明中提到：

> 為了保證檢索結果的準確和全面，必須根據當前考釋研究成果，對字形所歸屬的字頭進行細緻的認定。而相對於甲骨文、金文而言，戰國楚簡帛文字的發現比較遲，許多方面的研究尚不是很充分。特別是針對楚簡帛文字紛繁複雜的書寫異形和用字通假情況，許多字的考釋認定還沒有定論。這給以字為單位的字形檢索系統的開發帶來了很大的困難。[29]

因此，「戰國楚簡帛字形檢索系統」目前所收錄的資料，字形的選擇以形體清晰、結構具有代表性爲標準。目前該系統的編碼方式以GBK爲主，只提供GBK編碼

[26] 詳http://www.wenzi.cn/hanziziyuan/ZY_guwenzi.htm。
[27] 詳http://www.wenzi.cn/hanziziyuan/guwenzi/chuwenzi/chu_index.htm。
[28] 《戰國楚文獻檢索系統》見於 2003 年華東師範大學中國文字研究與應用中心開發的《戰國楚文字數位化處理系統》軟體光碟，該光碟由上海教育出版社出版。
[29] http://www.wenzi.cn/hanziziyuan/guwenzi/chuwenzi/chu_index.htm。

範圍內通用字頭的檢索，在編碼範圍內的字頭共有 1635 個，原始字形則有 4354 個。[30]該軟體除用單機光碟發售外，並開放一部份功能上傳到網站上供使用者使用：http://www.wenzi.cn/hanziziyuan/guwenzi/chuwenzi/chu_index.htm，查詢畫面如下：



目前網站僅提供「字頭」檢索功能。如輸入「羅」字，我們可以找到下列資料：

[30]據網頁內容的介紹，這個資料庫「下一步將不斷擴展檢索功能，並逐步發佈包括戰國楚簡帛文字所有字頭（包括Unicode擴展A、擴展B編碼範圍已收錄的楚文字隸定字形以及我們增補的楚文字隸定字形）、所有原始形體在內的檢索系統。」網址：http://www.wenzi.cn/hanziziyuan/guwenzi/chuwenzi/chu_index.htm。



根據上面的查詢結果，我們可以看到該網頁目前提供了字形「著錄名」、「篇名」、「著錄號」、「字形」等四個欄位資訊。在輸入「羅」字後我們找到了《包山楚簡》「文書 26」、《上海博物館藏戰國楚竹書》三「周易 56」、《上海博物館藏戰國楚竹書》四「柬大王泊旱 1」等三筆資料。

這個資料庫原則上提供了精確的查詢結果，一般使用者如果要找某字之楚簡文字，只需輸入某字之楷書字形，就可以找到相對應的楚簡字形。可惜的是，該網頁只提供了 1635 個楚簡字頭，裡面的圖檔雖然選自大量出土的楚國簡帛材料，可是全部的字形加起來也只有 4354 個。單就目前上博一至上博五約一萬七千多字的字形數量來看，該網站所提供的字形資料數量明顯過少。

柒、結論

編輯古文字編與建構古文字資料庫是一件極辛苦的工作，但許多的進階研究如果得到古文字編及相關資料庫的協助，必能減少研究者花費在檢索資料的龐大時間。研究者節省下來的時間便可以用來推進相關研究的深度、創造出更豐碩的研究成果。因此，無論是「瀚堂典藏」、「戰國楚簡帛文字典型形體檢索系統」，還是本資料庫《上博文字編》的建構與完成，共同目的都是希望研究者可以利用資料庫的便利性以輔助進行相關研究。在這個理念之下，筆者努力開發「上博電子文字編」資料庫，並提供「單字查詢與 Metadata 的呈現」、「各篇竹書內容分頁查詢」、「字頭查詢」、「上博簡電子文字編查詢」等幾種不同查詢功能。其中「上博簡電子文字編查詢」功能是目前較其他資料庫所勝出的「電子文字編」的建構

模式。筆者期待這種建構模式能夠帶給楚簡研究者更多的幫助，同時也希望使用者能給予更多的批評與指教，讓本資料庫、本文字編在建置及編輯上能更加周全完善！